

A REMARK ON THE LASSO AND THE DANTZIG SELECTOR

YOHANN DE CASTRO

ABSTRACT. Numerous authors have established a connection between the Compressed Sensing problem without noise and the estimation of the Gel'fand widths. This article shows that this connection is still true in the noisy case. Indeed, we investigate the lasso and the Dantzig selector in terms of the distortion of the design. This latter measures how far is the intersection between the kernel of the design matrix and the unit ℓ_1 -ball from an ℓ_2 -ball. In particular, we exhibit the weakest condition to get oracle inequalities in terms of the s -best term approximation.

1. INTRODUCTION

In the past decade much emphasis has been put on recovering a large number of unknown variables from few noisy observations. In particular, we consider the high-dimensional linear model where an experimenter observes a vector $y \in \mathbb{R}^n$ such that

$$y = X\beta^* + z,$$

where $X \in \mathbb{R}^{n \times p}$ is called the design matrix (known from the experimenter), $\beta^* \in \mathbb{R}^p$ is an unknown target vector one would like to recover, and $z \in \mathbb{R}^n$ is a stochastic error term that contains all the perturbations of the experiment. Assume that one can provide a constant $\lambda_n^0 \in \mathbb{R}$, as small as possible, such that

$$(1) \quad \lambda_n^0 \geq \|X^\top z\|_{\ell_\infty},$$

with an overwhelming probability (where $X^\top \in \mathbb{R}^{p \times n}$ denotes the transpose matrix of X). Observe that it is the only assumption on the noise throughout this paper. We recall a well-known result in the case where z is a n -multivariate Gaussian distribution.

Lemma 1.1 — Suppose that $z = (z_i)_{i=1}^n$ is such that the z_i 's are i.i.d with respect to a Gaussian distribution with mean zero and variance σ_n^2 . Choose $t \geq 1$ and set

$$\lambda_n^0(t) = (1+t) \cdot \|X\|_{\ell_{2,\infty}} \cdot \sigma_n \cdot \sqrt{\log p},$$

where $\|X\|_{\ell_{2,\infty}}$ denotes the maximum ℓ_2 -norm of the columns of X . Then,

$$\mathbb{P}(\lambda_n^0(t) \geq \|X^\top z\|_{\ell_\infty}) \geq 1 - \sqrt{2}/\left[(1+t)\sqrt{\pi \log p} p^{\frac{(1+t)^2}{2}-1}\right].$$

Motivated by recent issues in modern research areas, suppose that you have far less observation variables y_i than the unknown variables β_i^* , the so called $n \ll p$ setup. For instance, let us mention *Compressed Sensing* [Don06, CRT06] where one would like to simultaneously acquire and compress a signal using few (non-adaptive) linear measurements. In general terms, we are interested in accurately estimating the target vector β^* and/or the response $X\beta^*$ from few corrupted observations. During the past decade, this challenging issue has attracted a lot of attention among

Date: August 30, 2011.

Key words and phrases. Lasso, Oracle Inequalities, Compressed Sensing, Universal Distortion Property, Restricted Isoperimetric Property, Restricted Eigenvalue Condition, almost-Euclidean Sections of the ℓ_1 -ball, Distortion, Gelfand widths, Sparsity, Nullspace Property.

the statistical society. A breakthrough was initiated by R. Tibshirani in 1996 when he introduced the lasso [Tib96]. It is defined by

$$(2) \quad \beta^\ell \in \arg \min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2} \|X\beta - y\|_{\ell_2}^2 + \lambda_\ell \|\beta\|_{\ell_1} \right\},$$

where $\lambda_\ell > 0$ denotes a tuning parameter. Thirteen years later, this estimator continues to play a key role in our understanding of high-dimensional inverse problems. Its popularity might be due to the fact that this estimator is computationally tractable. Indeed, the lasso can be recasted in a Second Order Cone Program (SOCP) that can be solved using an interior point method. In the same way, E.J Candès and T. Tao [CT07] introduced the Dantzig selector as

$$(3) \quad \beta^d \in \arg \min_{\beta \in \mathbb{R}^p} \|\beta\|_{\ell_1} \text{ s.t. } \|X^\top(y - X\beta)\|_{\ell_\infty} \leq \lambda_d,$$

where $\lambda_d > 0$ is a tuning parameter. It is known that it can be recasted as a linear program. Hence, it is also computationally tractable. A great statistical challenge is then to find efficiently verifiable conditions on X ensuring that the lasso (2) and the Dantzig selector (3) would recover “most of the information” about the target vector β^* .

1.1. An oracle inequality. What do we precisely mean by “most of the information” about the target? What is the amount of information one could recover from few observations? That are two of the important questions raised by Compressed Sensing. Suppose that you want to find an s -sparse vector (i.e. a vector with at most s non-zero coefficients) that represents the target, then you would probably want that it contains the s largest (in magnitude) coefficients β_i^* . More precisely, denote $\mathcal{S}_* \subseteq \{1, \dots, p\}$ the set of the indices of the s largest coefficients. The s -best term approximation vector is $\beta_{\mathcal{S}_*}^* \in \mathbb{R}^p$ where $(\beta_{\mathcal{S}_*}^*)_i = \beta_i^*$ if $i \in \mathcal{S}_*$ and 0 otherwise. Observe that it is the s -sparse projection in respect to any ℓ_q -norm for $1 \leq q < +\infty$ (i.e. it minimizes the ℓ_q -distance to β^* among all the s -sparse vectors), and then the most natural approximation by an s -sparse vector.

Suppose that someone gives you all the keys to recover $\beta_{\mathcal{S}_*}^*$. More precisely, imagine that you know the subset \mathcal{S}_* a head of time in advance and that you observe $y^{\text{oracle}} = X\beta_{\mathcal{S}_*}^* + z$. This is an ideal situation referred as the oracle case. Assume that the noise z is a Gaussian white noise of standard deviation σ_n , i.e. $z \sim \mathcal{N}_n(0, \sigma_n^2 \text{Id}_n)$ where \mathcal{N}_n denotes the n -multivariate Gaussian distribution. Then the optimal estimator is the ordinary least square $\beta^{\text{ideal}} \in \mathbb{R}^p$ on the subset \mathcal{S}_* , namely

$$\beta^{\text{ideal}} \in \arg \min_{\substack{\beta \in \mathbb{R}^p \\ \text{Supp}(\beta) \subseteq \mathcal{S}_*}} \|X\beta - y^{\text{oracle}}\|_{\ell_2}^2,$$

where $\text{Supp}(\beta) \subseteq \{1, \dots, p\}$ denotes the support (i.e. the set of the indices of the non-zero coefficients) of the vector β . It holds

$$\|\beta^{\text{ideal}} - \beta^*\|_{\ell_1} = \|\beta^{\text{ideal}} - \beta_{\mathcal{S}_*}^*\|_{\ell_1} + \|\beta_{\mathcal{S}_*}^*\|_{\ell_1} \leq \sqrt{s} \|\beta^{\text{ideal}} - \beta_{\mathcal{S}_*}^*\|_{\ell_2} + \|\beta_{\mathcal{S}_*}^*\|_{\ell_1},$$

where $\beta_{\mathcal{S}_*^c}^* = \beta^* - \beta_{\mathcal{S}_*}^*$ denotes the ℓ_1 -error of the s -best term approximation. An easy calculation shows that

$$\mathbb{E} \|\beta^{\text{ideal}} - \beta_{\mathcal{S}_*}^*\|_{\ell_2}^2 = \text{Trace}((X_{\mathcal{S}_*}^\top X_{\mathcal{S}_*})^{-1}) \cdot \sigma_n^2 \geq \left(\frac{1}{\rho_1}\right)^2 \cdot \sigma_n^2 \cdot s,$$

where $X_{\mathcal{S}_*} \in \mathbb{R}^{n \times s}$ denotes the matrix composed by the columns $X_i \in \mathbb{R}^n$ of the matrix X such that $i \in \mathcal{S}_*$, and ρ_1 is the largest eigenvalue of X . It yields that

$$\left[\mathbb{E} \|\beta^{\text{ideal}} - \beta_{\mathcal{S}_*}^*\|_{\ell_1}^2 \right]^{1/2} \geq \frac{1}{\rho_1} \cdot \sigma_n \cdot \sqrt{s}.$$

In a nutshell, the ℓ_1 -distance between the target β^* and the optimal estimator β^{ideal} can be reasonably said of the order of

$$(4) \quad \frac{1}{\rho_1} \cdot \sigma_n \cdot s + \|\beta_{\mathcal{S}_*^c}^*\|_{\ell_1}.$$

In this article, we say that the lasso satisfies a *variable selection oracle inequality of order s* if and only if its ℓ_1 -distance to the target, namely $\|\beta^\ell - \beta^*\|_{\ell_1}$, is bounded by (4) up to a “satisfactory” multiplicative factor.

In some situations it could be interesting to have a good approximation of $X\beta^*$. In the oracle case, we have

$$\begin{aligned} \|X\beta^{ideal} - X\beta^*\|_{\ell_2} &\leq \|X\beta^{ideal} - X\beta_{\mathcal{S}_*}^*\|_{\ell_2} + \|X\beta_{\mathcal{S}_*^c}^*\|_{\ell_2}, \\ &\leq \|X\beta^{ideal} - X\beta_{\mathcal{S}_*}^*\|_{\ell_2} + \rho_1 \|\beta_{\mathcal{S}_*^c}^*\|_{\ell_1}. \end{aligned}$$

where ρ_1 denotes the largest singular value of X . An easy calculation gives that

$$\mathbb{E} \|X\beta^{ideal} - X\beta_{\mathcal{S}_*}^*\|_{\ell_2}^2 = \text{Trace}(X_{\mathcal{S}_*} (X_{\mathcal{S}_*}^\top X_{\mathcal{S}_*})^{-1} X_{\mathcal{S}_*}^\top) \cdot \sigma_n^2 = \sigma_n^2 \cdot s.$$

Hence a tolerable upper bound is given by

$$(5) \quad \sigma_n \cdot \sqrt{s} + \rho_1 \|\beta_{\mathcal{S}_*^c}^*\|_{\ell_1}.$$

We say that the lasso satisfies an *error prediction oracle inequality of order s* if and only if its prediction error is upper bounded by (5) up to a “satisfactory” multiplicative factor (say logarithmic in p).

1.2. The Universal Distortion Property (UDP). This article investigates a new sufficient condition to prove oracle inequalities for the lasso. We introduce the Universal Distortion Property $\text{UDP}(S_0, \kappa_0, \Delta)$ as follows.

Definition 1 ($\text{UDP}(S_0, \kappa_0, \Delta)$) — A matrix $X \in \mathbb{R}^{n \times p}$ satisfies the universal distortion condition of order S_0 , magnitude κ_0 and parameter Δ if and only if

- $1 \leq S_0 \leq p$,
- $0 < \kappa_0 < 1/2$,
- and for all $x \in \mathbb{R}^p$, for all integers $s \in \{1, \dots, S_0\}$, for all subsets $\mathcal{S} \subseteq \{1, \dots, p\}$ such that $|\mathcal{S}| = s$, it holds

$$(6) \quad \|x_{\mathcal{S}}\|_{\ell_1} \leq \Delta \sqrt{s} \|Xx\|_{\ell_2} + \kappa_0 \|x\|_{\ell_1}.$$

This property is similar to the Compatibility Condition of van de Geer and Bühlmann [vdGB09] although it is weaker (see Section 2 for a comparison with the usual conditions). As a matter of fact, every matrix satisfies the UDP condition with explicit parameters in terms of the geometry (e.g. the distortion) of its kernel, cf. Lemma (1.2).

1.2.1. The distortion. We recall the definition of the distortion.

Definition 2 — A subspace $\Gamma \subset \mathbb{R}^p$ has a distortion $1 \leq \delta \leq \sqrt{p}$ if and only if

$$\forall x \in \Gamma, \quad \|x\|_{\ell_1} \leq \sqrt{p} \|x\|_{\ell_2} \leq \delta \|x\|_{\ell_1}.$$

A long standing issue in approximation theory in Banach spaces is finding “almost-Euclidean” sections of the unit ℓ_1 -ball, i.e. subspaces with a distortion close to 1 and a dimension close to p . In particular, we recall that it has been established [Kas77] that there exists subspaces of dimension $p - n$ such that

$$(7) \quad \delta \leq C \left(\frac{p(1 + \log(p/n))}{n} \right)^{1/2}$$

where $C > 0$ is an universal constant. In other words, it was shown that, for all $n \leq p$, there exists a subspace Γ_n of dimension $p - n$ such that, for all $x \in \Gamma_n$,

$$\|x\|_{\ell_2} \leq C \left(\frac{1 + \log(p/n)}{n} \right)^{1/2} \|x\|_{\ell_1}.$$

We discuss recent deterministic constructions of almost sections of the ℓ_1 -ball in Section 3.

1.2.2. An universal property. Since that it is satisfied by all the full column rank matrices and that the parameters S_0 and Δ can be expressed in terms of the distortion, we name the property “Universal Distortion”. Indeed, we show the following lemma.

Lemma 1.2 — *Let $X \in \mathbb{R}^{n \times p}$ be a full column rank matrix. Denote δ the distortion of its kernel and ρ_n its smallest singular value. Let $0 < \kappa_0 < 1/2$ then X satisfies $\text{UDP}(S_0, \kappa_0, \Delta)$ where*

$$(8) \quad S_0 = \left(\frac{\kappa_0}{\delta} \right)^2 p \quad \text{and} \quad \Delta = \frac{2\delta}{\rho_n}.$$

This lemma is sharp in the following sense. The parameter S_0 represents (see Theorem 1.3) the maximum number of coefficients that can be recovered using lasso, we call it the *sparsity level*. It is known [CDD09] that the best bound one could expect is

$$S_{\text{opt}} \approx n / \log(p/n),$$

up to a multiplicative constant. In the case where (7) holds, the sparsity level satisfies

$$(9) \quad S_0 \approx \kappa_0^2 S_{\text{opt}}.$$

It shows that any design matrix with low distortion satisfies the UDP condition with an optimal sparsity level.

1.3. Results. The results presented here fold into two parts. In the first part we assume only that UDP holds. In particular, it is not exclude that one can get better upper bounds on the parameters than Lemma 1.2. As a matter of fact, the smaller Δ is the sharper the oracle inequalities are. In the second part, we give oracle inequalities in terms of only the distortion of the design.

Theorem 1.3 — *Let $X \in \mathbb{R}^{n \times p}$ be a full column rank matrix. Assume that X satisfies $\text{UDP}(S_0, \kappa_0, \Delta)$ and that (1) holds. Then for any*

$$(10) \quad \lambda_\ell > \lambda_n^0 / (1 - 2\kappa_0),$$

it holds

$$(11) \quad \|\beta^\ell - \beta^*\|_{\ell_1} \leq \frac{2}{\left(1 - \frac{\lambda_n^0}{\lambda_\ell}\right) - 2\kappa_0} \min_{\substack{S \subseteq \{1, \dots, p\}, \\ |S|=s, s \leq S_0}} \left(\lambda_\ell \Delta^2 s + \|\beta_{S^c}^*\|_{\ell_1} \right).$$

Invoking Lemma 1.2, the following holds: For every full column rank matrix $X \in \mathbb{R}^{n \times p}$, for all $0 < \kappa_0 < 1/2$ and λ_ℓ satisfying (10), we have

$$(12) \quad \|\beta^\ell - \beta^*\|_{\ell_1} \leq \frac{2}{\left(1 - \frac{\lambda_n^0}{\lambda_\ell}\right) - 2\kappa_0} \min_{\substack{S \subseteq \{1, \dots, p\}, \\ |S|=s, \\ s \leq (\kappa_0/\delta)^2 p}} \left(\lambda_\ell \cdot \frac{4\delta^2}{\rho_n^2} \cdot s + \|\beta_{S^c}^*\|_{\ell_1} \right),$$

where ρ_n denotes the smallest eigenvalue of X and δ the distortion of its kernel.

◇ Consider the case where the noise satisfies the hypothesis of Lemma 1.1 and take $\lambda_n^0 = \lambda_n^0(1)$. Assume that κ_0 is constant (say $\kappa_0 = 1/3$) and take $\lambda_\ell = 3\lambda_n^0$ then (11) becomes

$$\|\beta^\ell - \beta^*\|_{\ell_1} \leq 12 \min_{\substack{\mathcal{S} \subseteq \{1, \dots, p\}, \\ |\mathcal{S}|=s, s \leq S_0.}} \left(6 \|X\|_{\ell_2, \infty} \cdot \Delta^2 \sqrt{\log p} \cdot \sigma_n s + \|\beta_{\mathcal{S}^c}^*\|_{\ell_1} \right),$$

which is an oracle inequality up to a multiplicative factor $\Delta^2 \sqrt{\log p}$. In the same way, (12) becomes

$$\|\beta^\ell - \beta^*\|_{\ell_1} \leq 12 \min_{\substack{\mathcal{S} \subseteq \{1, \dots, p\}, \\ |\mathcal{S}|=s, \\ s \leq p/9\delta^2.}} \left(24 \|X\|_{\ell_2, \infty} \cdot \frac{\delta^2 \sqrt{\log p}}{\rho_n} \cdot \frac{1}{\rho_n} \sigma_n s + \|\beta_{\mathcal{S}^c}^*\|_{\ell_1} \right),$$

which is an oracle inequality up to a multiplicative factor $C_{mult} := (\delta^2 \sqrt{\log p})/\rho_n$.

◇ In the optimal case (7), this latter becomes:

$$(13) \quad C_{mult} = C \cdot \frac{p(1 + \log(p/n)) \sqrt{\log p}}{n \rho_n},$$

where $C > 0$ is the same universal constant as in (7). Roughly speaking, up to a factor of the order of (13), the lasso is as good as the oracle that knows the S_0 -best term approximation of the target. Moreover, as mentioned in (9), S_0 is an optimal sparsity level. However, this multiplicative constant takes small values for a restrictive range of the parameter n . As a matter of fact, it is meaningful when n is a constant fraction of p .

Similarly, we shows oracle inequalities in error prediction in terms of the distortion of the kernel of the design.

Theorem 1.4 — *Let $X \in \mathbb{R}^{n \times p}$ be a full column rank matrix. Assume that X satisfies $\text{UDP}(S_0, \kappa_0, \Delta)$ and that (1) holds. Then for any*

$$(10) \quad \lambda_\ell > \lambda_n^0 / (1 - 2\kappa_0),$$

it holds

$$(14) \quad \|X\beta^\ell - X\beta^*\|_{\ell_2} \leq \min_{\substack{\mathcal{S} \subseteq \{1, \dots, p\}, \\ |\mathcal{S}|=s, s \leq S_0.}} \left[4\lambda_\ell \Delta \sqrt{s} + \frac{\|\beta_{\mathcal{S}^c}^*\|_{\ell_1}}{\Delta \sqrt{s}} \right].$$

Invoking Lemma 1.2, the following holds: For every full column rank matrix $X \in \mathbb{R}^{n \times p}$, for all $0 < \kappa_0 < 1/2$ and λ_ℓ satisfying (10), we have

$$(15) \quad \|X\beta^\ell - X\beta^*\|_{\ell_2} \leq \min_{\substack{\mathcal{S} \subseteq \{1, \dots, p\}, \\ |\mathcal{S}|=s, \\ s \leq (\kappa_0/\delta)^2 p.}} \left[4\lambda_\ell \cdot \frac{2\delta}{\rho_n} \cdot \sqrt{s} + \frac{1}{2\delta\sqrt{s}} \cdot \rho_n \|\beta_{\mathcal{S}^c}^*\|_{\ell_1} \right],$$

where ρ_n denotes the smallest eigenvalue of X and δ the distortion of its kernel.

◇ Consider the case where the noise satisfies the hypothesis of Lemma 1.1 and take $\lambda_n^0 = \lambda_n^0(1)$. Assume that κ_0 is constant (say $\kappa_0 = 1/3$) and take $\lambda_\ell = 3\lambda_n^0$ then (14) becomes

$$\|X\beta^\ell - X\beta^*\|_{\ell_2} \leq \min_{\substack{\mathcal{S} \subseteq \{1, \dots, p\}, \\ |\mathcal{S}|=s, s \leq S_0.}} \left[24 \|X\|_{\ell_2, \infty} \cdot \Delta \sqrt{\log p} \cdot \sigma_n \sqrt{s} + \frac{\|\beta_{\mathcal{S}^c}^*\|_{\ell_1}}{\Delta \sqrt{s}} \right],$$

which is not an oracle inequality *stricto sensu* because of $1/(\Delta \sqrt{s})$ in the second term. As a matter of fact, it tends to lower the s -best term approximation term

$\|\beta_{S^c}^*\|_{\ell_1}$. Nevertheless, it is “almost” an oracle inequality up to a multiplicative factor of the order of $\Delta\sqrt{\log p}$. In the same way, (15) becomes

$$\|X\beta^\ell - X\beta^*\|_{\ell_2} \leq \min_{\substack{S \subseteq \{1, \dots, p\}, \\ |S|=s, \\ s \leq p/9\delta^2}} \left[48 \|X\|_{\ell_2, \infty} \cdot \frac{\delta\sqrt{\log p}}{\rho_n} \cdot \frac{1}{\rho_n} \sigma_n \sqrt{s} + \frac{1}{2\delta\sqrt{s}} \cdot \rho_n \|\beta_{S^c}^*\|_{\ell_1} \right],$$

which is an oracle inequality up to a multiplicative factor $C'_{mult} := (\delta\sqrt{\log p})/\rho_n$.

◇ In the optimal case (7), this latter becomes:

$$(16) \quad C'_{mult} = C \cdot \frac{(p \log p (1 + \log(p/n)))^{1/2}}{\rho_n \sqrt{n}},$$

where $C > 0$ is the same universal constant as in (7).

1.4. Results for the Dantzig selector. Similarly, we derive the same results for the Dantzig selector. The only difference is that the parameter κ_0 must be less than $1/4$. Here again the results folds into two parts. In the first one, we only assume that UDP holds. In the second, we invoke Lemma 1.2 to derive results in terms of the distortion of the design.

Theorem 1.5 — *Let $X \in \mathbb{R}^{n \times p}$ be a full column rank matrix. Assume that X satisfies $\text{UDP}(S_0, \kappa_0, \Delta)$ with $\kappa_0 < 1/4$ and that (1) holds. Then for any*

$$(17) \quad \lambda_d > \lambda^0 / (1 - 4\kappa_0),$$

it holds

$$(18) \quad \|\beta^d - \beta^*\|_{\ell_1} \leq \frac{4}{\left(1 - \frac{\lambda^0}{\lambda_d}\right) - 4\kappa_0} \min_{\substack{S \subseteq \{1, \dots, p\}, \\ |S|=s, s \leq S_0}} \left(\lambda_d \Delta^2 s + \|\beta_{S^c}^*\|_{\ell_1} \right).$$

Invoking Lemma 1.2, the following holds: For every full column rank matrix $X \in \mathbb{R}^{n \times p}$, for all $0 < \kappa_0 < 1/4$ and λ_d satisfying (17), we have

$$(19) \quad \|\beta^d - \beta^*\|_{\ell_1} \leq \frac{4}{\left(1 - \frac{\lambda^0}{\lambda_d}\right) - 4\kappa_0} \min_{\substack{S \subseteq \{1, \dots, p\}, \\ |S|=s, \\ s \leq (\kappa_0/\delta)^2 p}} \left(\lambda_d \cdot \frac{4\delta^2}{\rho_n^2} \cdot s + \|\beta_{S^c}^*\|_{\ell_1} \right),$$

where ρ_n denotes the smallest eigenvalue of X and δ the distortion of its kernel.

The prediction error is given by the following theorem.

Theorem 1.6 — *Let $X \in \mathbb{R}^{n \times p}$ be a full column rank matrix. Assume that X satisfies $\text{UDP}(S_0, \kappa_0, \Delta)$ with $\kappa_0 < 1/4$ and that (1) holds. Then for any*

$$(17) \quad \lambda_d > \lambda^0 / (1 - 4\kappa_0),$$

it holds

$$(20) \quad \|X\beta^d - X\beta^*\|_{\ell_2} \leq \min_{\substack{S \subseteq \{1, \dots, p\}, \\ |S|=s, s \leq S_0}} \left[4\lambda_d \Delta \sqrt{s} + \frac{\|\beta_{S^c}^*\|_{\ell_1}}{\Delta \sqrt{s}} \right].$$

Invoking Lemma 1.2, the following holds: For every full column rank matrix $X \in \mathbb{R}^{n \times p}$, for all $0 < \kappa_0 < 1/4$ and λ_d satisfying (10), we have

$$(21) \quad \|X\beta^d - X\beta^*\|_{\ell_2} \leq \min_{\substack{S \subseteq \{1, \dots, p\}, \\ |S|=s, \\ s \leq (\kappa_0/\delta)^2 p}} \left[4\lambda_d \cdot \frac{2\delta}{\rho_n} \cdot \sqrt{s} + \frac{1}{2\delta\sqrt{s}} \cdot \rho_n \|\beta_{S^c}^*\|_{\ell_1} \right],$$

where ρ_n denotes the smallest eigenvalue of X and δ the distortion of its kernel.

Observe that the same comments as in the lasso case (e.g. (13), (16)) hold. Eventually, every result in constructing deterministic almost-Euclidean sections gives design that satisfies the oracle inequalities above.

1.5. Organization of the paper. The paper folds into three parts. The next section shows that UDP is the weakest condition on the lasso to have oracle inequalities. In particular we show that RIP implies UDP. Section 3 recalls the best known results for constructing subspaces with low-distortion. The last section is devoted to the proofs of the results.

2. AN OVERVIEW OF THE STANDARDS CONDITIONS

Oracle inequalities for the lasso have been established under a variety of different conditions on the design. An remarkable overview can be found in the article of van de Geer and Bühlmann. We recall some sufficient conditions here. For all $s \in \{1, \dots, p\}$, we denote by $\Sigma_s \subseteq \mathbb{R}^p$ the set of all the s -sparse vectors.

- ◆ **Restricted Isoperimetric Property:** A matrix $X \in \mathbb{R}^{n \times p}$ satisfies $RIP(\theta_S)$ if and only if there exists $0 < \theta_S < 1$ (as small as possible) such that for all $s \in \{1, \dots, S\}$, for all $\forall \gamma \in \Sigma_s$, it holds

$$(1 - \theta_S) \|\gamma\|_{\ell_2}^2 \leq \|X\gamma\|_{\ell_2}^2 \leq (1 + \theta_S) \|\gamma\|_{\ell_2}^2.$$

The constant θ_S is called the S -restricted isometry constant.

- ◆ **Restricted Eigenvalue Assumption [BRT09]:** A matrix $X \in \mathbb{R}^{n \times p}$ satisfies $RE(S, c_0)$ if and only if

$$\kappa(S, c_0) = \min_{\substack{S \subseteq \{1, \dots, p\} \\ |S| \leq S}} \min_{\substack{\gamma \neq 0 \\ \|\gamma_{S^c}\|_{\ell_1} \leq c_0 \|\gamma_S\|_{\ell_1}}} \frac{\|X\gamma\|_{\ell_2}}{\|\gamma_S\|_{\ell_2}} > 0.$$

The constant $\kappa(S, c_0)$ is called the (S, c_0) -restricted ℓ_2 -eigenvalue.

- ◆ **Compatibility Condition [vdGB09]:** A matrix $X \in \mathbb{R}^{n \times p}$ satisfies the condition $Compatibility(S, c_0)$ if and only if

$$\phi(S, c_0) = \min_{\substack{S \subseteq \{1, \dots, p\} \\ |S| \leq S}} \min_{\substack{\gamma \neq 0 \\ \|\gamma_{S^c}\|_{\ell_1} \leq c_0 \|\gamma_S\|_{\ell_1}} \frac{\sqrt{|S|} \|X\gamma\|_{\ell_2}}{\|\gamma_S\|_{\ell_1}} > 0.$$

The constant $\phi(S, c_0)$ is called the (S, c_0) -restricted ℓ_1 -eigenvalue.

- ◆ **$H_{S,1}$ Condition [JN10]:** $X \in \mathbb{R}^{n \times p}$ satisfies the $H_{S,1}(\kappa)$ condition (with $\kappa < 1/2$) if and only if for all $\gamma \in \mathbb{R}^p$ and for all $S \subseteq \{1, \dots, p\}$ such that $|S| \leq S$, it holds

$$(22) \quad \|\gamma_S\|_{\ell_1} \leq \hat{\lambda} S \|X\gamma\|_{\ell_2} + \kappa \|\gamma\|_{\ell_1},$$

where $\hat{\lambda}$ denotes the maximum of the ℓ_2 -norms of the columns in X .

Remark. This latter condition is weaker than the UDP condition nevertheless the authors [JN10] established limits of performance on their conditions: the condition $H_{s,\infty}(1/3)$ (that implies $H_{s,1}(1/3)$) is feasible only in a severe restricted range of the sparsity parameter s . Notice that this is not the case of the UDP condition, the equality (9) shows that it is feasible for a large range of the sparsity parameter s (indeed an optimal range, cf. (9)).

Let us emphasize that the above description is not meant to be exhaustive. In particular we do not mention the irrepresentable condition [ZY06] which ensures exact recovery of the support. The next proposition shows that the UDP condition is weaker than RIP, RE and Compatibility condition.

Proposition 2.1 — *Let $X \in \mathbb{R}^{n \times p}$ be a full column rank matrix, then the following is true:*

- ◆ *The $\text{RIP}(\theta_{5S})$ condition with $\theta_{5S} < \sqrt{2} - 1$ implies $\text{UDP}(S, \kappa_0, \Delta)$ for all pairs (κ_0, Δ) such that*

$$(23) \quad \left[1 + 2 \left[\frac{1 - \theta_{5S}}{1 + \theta_{5S}} \right]^{\frac{1}{2}} \right]^{-1} < \kappa_0 < \frac{1}{2}, \text{ and } \Delta \geq \left[\sqrt{1 - \theta_{5S}} + \frac{\kappa_0 - 1}{2\kappa_0} \sqrt{1 + \theta_{5S}} \right]^{-1}.$$

- ◆ *The $\text{RE}(S, c_0)$ condition implies $\text{UDP}(S, c_0, \kappa(S, c_0)^{-1})$.*
- ◆ *The Compatibility(S, c_0) condition implies $\text{UDP}(S, c_0, \phi(S, c_0)^{-1})$.*

Proof. It is obvious that $\text{RE}(S, c_0)$ condition implies $\text{UDP}(S, c_0, \kappa(S, c_0)^{-1})$, and that $\text{Compatibility}(S, c_0)$ condition implies $\text{UDP}(S, c_0, \phi(S, c_0)^{-1})$.

Assume that X satisfies $\text{RIP}(\theta_{5S})$. Let $\gamma \in \mathbb{R}^p$, $s \in \{1, \dots, S_0\}$, and $T_0 \subseteq \{1, \dots, p\}$ such that $|T_0| = s$. Choose a pair (κ_0, Δ) as in (23).

◆ If $\|\gamma_{T_0}\|_{\ell_1} \leq \kappa_0 \|\gamma\|_{\ell_1}$ then $\|\gamma_{T_0}\|_{\ell_1} \leq \Delta \sqrt{s} \|X\gamma\|_{\ell_2} + \kappa_0 \|\gamma\|_{\ell_1}$.

◆ Suppose that $\|\gamma_{T_0}\|_{\ell_1} > \kappa_0 \|\gamma\|_{\ell_1}$ then

$$(24) \quad \|\gamma_{T_0^c}\|_{\ell_1} < \frac{1 - \kappa_0}{\kappa_0} \|\gamma_{T_0}\|_{\ell_1}.$$

Denote T_1 the set of the indices of the $4s$ largest coefficients (in absolute value) in T_0^c , denote T_2 the set of the indices of the $4s$ largest coefficients in $(T_0 \cup T_1)^c$, etc... Hence we decompose T_0^c into disjoint sets

$$T_0^c = T_1 \cup T_2 \cup \dots \cup T_l.$$

Using (24), it yields

$$(25) \quad \sum_{i \geq 2} \|\gamma_{T_i}\|_{\ell_2} \leq (4s)^{-1/2} \sum_{i \geq 1} \|\gamma_{T_i}\|_{\ell_1} = (4s)^{-1/2} \|\gamma_{T_0^c}\|_{\ell_1} \leq \frac{1 - \kappa_0}{2\kappa_0 \sqrt{s}} \|\gamma_{T_0}\|_{\ell_1}$$

Using $\text{RIP}(\theta_{5S})$ and (25), it follows that

$$\begin{aligned} \|X\gamma\|_{\ell_2} &\geq \|X(\gamma_{(T_0 \cup T_1)})\|_{\ell_2} - \sum_{i \geq 2} \|X(\gamma_{T_i})\|_{\ell_2}, \\ &\geq \sqrt{1 - \theta_{5S}} \|\gamma_{(T_0 \cup T_1)}\|_{\ell_2} - \sqrt{1 + \theta_{5S}} \sum_{i \geq 2} \|\gamma_{T_i}\|_{\ell_2}, \\ &\geq \sqrt{1 - \theta_{5S}} \|\gamma_{T_0}\|_{\ell_2} - \sqrt{1 + \theta_{5S}} \frac{1 - \kappa_0}{2\kappa_0} \frac{\|\gamma_{T_0}\|_{\ell_1}}{\sqrt{s}}, \\ &\geq \left[\sqrt{1 - \theta_{5S}} + \frac{\kappa_0 - 1}{2\kappa_0} \sqrt{1 + \theta_{5S}} \right] \frac{\|\gamma_{T_0}\|_{\ell_1}}{\sqrt{s}}, \\ &= \frac{\sqrt{1 + \theta_{5S}}}{2\kappa_0} \left[1 + 2 \left(\frac{1 - \theta_{5S}}{1 + \theta_{5S}} \right)^{\frac{1}{2}} \right] \left[\kappa_0 - \left[1 + 2 \left(\frac{1 - \theta_{5S}}{1 + \theta_{5S}} \right)^{\frac{1}{2}} \right]^{-1} \right] \frac{\|\gamma_{T_0}\|_{\ell_1}}{\sqrt{s}}. \end{aligned}$$

The lower bound on κ_0 shows that the right hand side is positive. Observe that we took Δ such that this latter is exactly $\|\gamma_{T_0}\|_{\ell_1} / (\Delta \sqrt{s})$. Eventually, we get

$$\|\gamma_{T_0}\|_{\ell_1} \leq \Delta \sqrt{s} \|X\gamma\|_{\ell_2} \leq \Delta \sqrt{s} \|X\gamma\|_{\ell_2} + \kappa_0 \|\gamma\|_{\ell_1}.$$

This ends the proofs. \square

The UDP condition is the weakest condition among all the condition on the lasso.

3. THE DISTORTION OF THE DESIGN

One of the big issue in modern statistics is to find verifiable conditions. This question is valuable to the statistics community since one knows that the RIP condition (which is the key stone of Compressed Sensing) cannot be computationally checked for a given matrix. To overcome this difficulty, at the price of weaker results, we investigate the role of the distortion in high-dimensional regression. It is known that there is a connection between the Compressed Sensing problem and the problem of estimating the distortion. This framework was studied by numerous authors [CDD09, KT07, DeV07] and might interest both people working on building deterministic almost Euclidean section of the ℓ_1 -ball and those looking for deterministic design for Compressed Sensing. Table 1 presents some important results for constructing almost-Euclidean sections of the ℓ_1 -ball. The last line of Table 1 deals with the optimal case derived from a probabilistic construction. Even though this construction has been established in the late '70 there is no deterministic proof of it.

Reference	Distortion	Co-dimension	Randomness
[Ind07]	$1 + \varepsilon$	$p - p^{1-o_\varepsilon(1)}$	Explicit
[GLR08]	$\log(p)^{O_\eta(\log \log \log p)}$	ηp	Explicit
[IS10]	$1 + \varepsilon$	$(1 - (\gamma\varepsilon)^{O(1/\gamma)})p$	$O(p^\gamma)$
[Kas77]	$C(p(1 + \log(p/n))/n)^{1/2}$	n	np

TABLE 1. The best known results for constructing almost Euclidean subspaces. The parameters $\varepsilon, \eta, \gamma \in (0, 1)$ are assumed to be constants, although the dependence on them is subsumed by the big-Oh notation. The parameter $C > 0$ denotes an universal constant.

Most of the explicit constructions can be viewed as related to the context of error-correcting codes. Indeed, the construction of [Ind07] is based on amplifying the minimum distance of a code using expanders. While the construction of [GLR08] is based on Low-Density Parity Check (LDPC) codes. Lastly, the construction of [IS10] is related to the tensor product of error-correcting codes. The main reason of this surprising fact is that the vectors of a subspace of low distortion must be “well-spread”, i.e. a small subset of its coordinates cannot contain most of its ℓ_2 -norm (cf [Ind07, GLR08]). This property is required from a good error-correcting code, where the weight (i.e. the ℓ_0 -norm) of each codeword cannot be concentrated on a small subset of its coordinates. Similarly, this property was intensively studied in Compressed Sensing, see for instance the Nullspace Property in [CDD09].

4. PROOFS

Proof of Lemma 1.1 — Observe that $X^\top z \sim \mathcal{N}_p(0, \sigma_n^2 X^\top X)$. Hence,

$$\forall j = 1, \dots, p, \quad X_j^\top z \sim \mathcal{N}(0, \sigma_n^2 \|X_j\|_{\ell_2}^2).$$

Using Šidák’s inequality [Š68], it yields

$$\mathbb{P}(\|X^\top z\|_{\ell_\infty} \leq \lambda_n^0) \geq \mathbb{P}(\|\tilde{z}\|_{\ell_\infty} \leq \lambda_n^0) = \prod_{i=1}^p \mathbb{P}(|\tilde{z}_i| \leq \lambda_n^0),$$

where the \tilde{z}_i ’s are i.i.d. with respect to $\mathcal{N}(0, \sigma_n^2 \|X\|_{\ell_2, \infty}^2)$. Denote Φ and φ respectively the cumulative distribution function and the probability density function of

the standard normal. Set $\theta = (1+t)\sqrt{\log p}$. It holds

$$\prod_{i=1}^p \mathbb{P}(|\tilde{z}_i| \leq \lambda_n^0) = \mathbb{P}(|z_1| \leq \lambda_n^0)^p = (2\Phi(\theta) - 1)^p > (1 - 2\varphi(\theta)/\theta)^p,$$

using an integration by parts to get $1 - \Phi(\theta) < \varphi(\theta)/\theta$. It yields that

$$\mathbb{P}(\|X^\top z\|_{\ell_\infty} \leq \lambda_n^0) \geq (1 - 2\varphi(\theta)/\theta)^p \geq 1 - 2p \frac{\varphi(\theta)}{\theta} = 1 - \frac{\sqrt{2}}{(1+t)\sqrt{\pi \log p} p^{\frac{(1+t)^2}{2}-1}}.$$

This concludes the proof. \square

Proof of Lemma 1.2 — Consider the following singular value decomposition $X = U^\top D A$ where

- $\diamond U \in \mathbb{R}^{n \times n}$ is such that $U U^\top = \text{Id}_n$,
- $\diamond D = \text{Diag}(\rho_1, \dots, \rho_n)$ is a diagonal matrix where $\rho_1 \geq \dots \geq \rho_n > 0$ are the singular values of X ,
- \diamond and $A \in \mathbb{R}^{n \times p}$ is such that $A A^\top = \text{Id}_n$.

We recall that the only assumption on the design is that it has full column rank which yields that $\rho_n > 0$. Let δ be the distortion of the kernel Γ of the design. Denote by π_Γ (resp. π_{Γ^\perp}) the ℓ_2 -projection onto Γ (resp. Γ^\perp). Let $\gamma \in \mathbb{R}^p$, then $\gamma = \pi_\Gamma(\gamma) + \pi_{\Gamma^\perp}(\gamma)$. An easy calculation shows that

$$\pi_{\Gamma^\perp}(\gamma) = A^\top A \gamma.$$

Let $s \in \{1, \dots, S\}$ and let $\mathcal{S} \subseteq \{1, \dots, p\}$ be such that $|\mathcal{S}| = s$. It holds,

$$\begin{aligned} \|\gamma_{\mathcal{S}}\|_{\ell_1} &\leq \sqrt{s} \|\gamma\|_{\ell_2}, \\ &= \sqrt{s} \|\pi_\Gamma(\gamma)\|_{\ell_2} + \sqrt{s} \|\pi_{\Gamma^\perp}(\gamma)\|_{\ell_2}, \\ &\leq \frac{\sqrt{s}}{\sqrt{p}} \delta \|\pi_\Gamma(\gamma)\|_{\ell_1} + \sqrt{s} \|A^\top A \gamma\|_{\ell_2}, \\ &\leq \frac{\sqrt{s}}{\sqrt{p}} \delta (\|\gamma\|_{\ell_1} + \|\pi_{\Gamma^\perp}(\gamma)\|_{\ell_1}) + \sqrt{s} \|A \gamma\|_{\ell_2}, \\ &\leq \frac{\sqrt{s}}{\sqrt{p}} \delta \|\gamma\|_{\ell_1} + \delta \sqrt{s} \|A^\top A \gamma\|_{\ell_2} + \sqrt{s} \|A \gamma\|_{\ell_2}, \\ &\leq \frac{\sqrt{s}}{\sqrt{p}} \delta \|\gamma\|_{\ell_1} + (1 + \delta) \sqrt{s} \|A \gamma\|_{\ell_2}, \\ &\leq \frac{\sqrt{s}}{\sqrt{p}} \delta \|\gamma\|_{\ell_1} + \frac{1 + \delta}{\rho_n} \sqrt{s} \|X \gamma\|_{\ell_2}, \\ &\leq \frac{\sqrt{s}}{\sqrt{p}} \delta \|\gamma\|_{\ell_1} + \frac{2\delta}{\rho_n} \sqrt{s} \|X \gamma\|_{\ell_2}, \end{aligned}$$

using the triangular inequality and the distortion of the kernel Γ . Eventually, set $\kappa_0 = (\sqrt{S}/\sqrt{p}) \delta$ and $\Delta = 2\delta/\rho_n$. This ends the proof. \square

Proof of Theorem 1.3 — We recall that λ_n^0 denotes an upper bound on the amplification of the noise, see (1). We begin with a standard result.

Lemma 4.1 — Let $h = \beta^\ell - \beta^\star \in \mathbb{R}^p$ and $\lambda_\ell \geq \lambda_n^0$. Then, for all subsets $\mathcal{S} \subseteq \{1, \dots, p\}$, it holds,

$$(26) \quad \frac{1}{2\lambda_\ell} \left[\frac{1}{2} \|X h\|_{\ell_2}^2 + (\lambda_\ell - \lambda_n^0) \|h\|_{\ell_1} \right] \leq \|h_{\mathcal{S}}\|_{\ell_1} + \|\beta_{\mathcal{S}^c}^\star\|_{\ell_1}.$$

Proof. By optimality, we have

$$\frac{1}{2}\|X\beta^\ell - y\|_{\ell_2}^2 + \lambda_\ell \|\beta^\ell\|_{\ell_1} \leq \frac{1}{2}\|X\beta^* - y\|_{\ell_2}^2 + \lambda_\ell \|\beta^*\|_{\ell_1}.$$

It yields

$$\frac{1}{2}\|Xh\|_{\ell_2}^2 - \langle X^\top z, h \rangle + \lambda_\ell \|\beta^\ell\|_{\ell_1} \leq \lambda_\ell \|\beta^*\|_{\ell_1}.$$

Let $\mathcal{S} \subseteq \{1, \dots, p\}$, we have

$$\begin{aligned} \frac{1}{2}\|Xh\|_{\ell_2}^2 + \lambda_\ell \|\beta_{\mathcal{S}^c}^\ell\|_{\ell_1} &\leq \lambda_\ell (\|\beta_{\mathcal{S}}^*\|_{\ell_1} - \|\beta_{\mathcal{S}}^\ell\|_{\ell_1}) + \lambda_\ell \|\beta_{\mathcal{S}^c}^*\|_{\ell_1} + \langle X^\top z, h \rangle, \\ &\leq \lambda_\ell \|h_{\mathcal{S}}\|_{\ell_1} + \lambda_\ell \|\beta_{\mathcal{S}^c}^*\|_{\ell_1} + \lambda_n^0 \|h\|_{\ell_1}, \end{aligned}$$

using (1). Adding $\lambda_\ell \|\beta_{\mathcal{S}^c}^*\|_{\ell_1}$ on both sides, it holds

$$\frac{1}{2}\|Xh\|_{\ell_2}^2 + (\lambda_\ell - \lambda_n^0) \|h_{\mathcal{S}^c}\|_{\ell_1} \leq (\lambda_\ell + \lambda_n^0) \|h_{\mathcal{S}}\|_{\ell_1} + 2\lambda_\ell \|\beta_{\mathcal{S}^c}^*\|_{\ell_1}.$$

Adding $(\lambda_\ell - \lambda_n^0) \|h_{\mathcal{S}}\|_{\ell_1}$ on both sides, we conclude the proof. \square

Using (6) and (26), it follows that

$$(27) \quad \frac{1}{2\lambda_\ell} \left[\frac{1}{2}\|Xh\|_{\ell_2}^2 + (\lambda_\ell - \lambda_n^0) \|h\|_{\ell_1} \right] \leq \Delta\sqrt{s} \|Xh\|_{\ell_2} + \kappa_0 \|h\|_{\ell_1} + \|\beta_{\mathcal{S}^c}^*\|_{\ell_1}.$$

It yields,

$$\begin{aligned} \left[\frac{1}{2} \left(1 - \frac{\lambda_n^0}{\lambda_\ell} \right) - \kappa_0 \right] \|h\|_{\ell_1} &\leq \left(-\frac{1}{4\lambda_\ell} \|Xh\|_{\ell_2}^2 + \Delta\sqrt{s} \|Xh\|_{\ell_2} \right) + \|\beta_{\mathcal{S}^c}^*\|_{\ell_1}, \\ &\leq \lambda_\ell \Delta^2 s + \|\beta_{\mathcal{S}^c}^*\|_{\ell_1}, \end{aligned}$$

using the fact that the polynomial $x \mapsto -(1/4\lambda_\ell)x^2 + \Delta\sqrt{s}x$ is not greater than $\lambda_\ell \Delta^2 s$. This concludes the proof. \square

Proof of Theorem 1.5 — We begin with a standard result.

Lemma 4.2 — Let $h = \beta^\ell - \beta^* \in \mathbb{R}^p$ and $\lambda_\ell \geq \lambda_n^0$. Then, for all subsets $\mathcal{S} \subseteq \{1, \dots, p\}$, it holds,

$$(28) \quad \frac{1}{4\lambda_d} \left[\|Xh\|_{\ell_2}^2 + (\lambda_d - \lambda_n^0) \|h\|_{\ell_1} \right] \leq \|h_{\mathcal{S}}\|_{\ell_1} + \|\beta_{\mathcal{S}^c}^*\|_{\ell_1}.$$

Proof. Set $h = \beta^* - \beta^d$. Recall that $\|X^\top z\|_{\ell_\infty} \leq \lambda_n^0$, it yields

$$\begin{aligned} \|Xh\|_{\ell_2}^2 &\leq \|X^\top Xh\|_{\ell_\infty} \|h\|_{\ell_1} \\ &= \|X^\top (y - X\beta^d) + X^\top (X\beta^* - y)\|_{\ell_\infty} \|h\|_{\ell_1} \\ &\leq (\lambda_d + \lambda_n^0) \|h\|_{\ell_1}. \end{aligned}$$

Hence we get

$$(29) \quad \|Xh\|_{\ell_2}^2 - (\lambda_d + \lambda_n^0) \|h_{\mathcal{S}^c}\|_{\ell_1} \leq (\lambda_d + \lambda_n^0) \|h_{\mathcal{S}}\|_{\ell_1}.$$

Since β^* is feasible, it yields $\|\beta^d\|_{\ell_1} \leq \|\beta^*\|_{\ell_1}$. Thus,

$$\|\beta_{\mathcal{S}^c}^d\|_{\ell_1} \leq (\|\beta_{\mathcal{S}}^*\|_{\ell_1} - \|\beta_{\mathcal{S}}^d\|_{\ell_1}) + \|\beta_{\mathcal{S}^c}^*\|_{\ell_1} \leq \|h_{\mathcal{S}}\|_{\ell_1} + \|\beta_{\mathcal{S}^c}^*\|_{\ell_1}.$$

Since $\|h_{\mathcal{S}^c}\|_{\ell_1} \leq \|\beta_{\mathcal{S}^c}^d\|_{\ell_1} + \|\beta_{\mathcal{S}^c}^*\|_{\ell_1}$, it yields

$$(30) \quad \|h_{\mathcal{S}^c}\|_{\ell_1} \leq \|h_{\mathcal{S}}\|_{\ell_1} + 2\|\beta_{\mathcal{S}^c}^*\|_{\ell_1}.$$

Combining (29) + $2\lambda_d \cdot (30)$, we get

$$\|Xh\|_{\ell_2}^2 + (\lambda_d - \lambda_n^0) \|h_{\mathcal{S}^c}\|_{\ell_1} \leq (3\lambda_d + \lambda_n^0) \|h_{\mathcal{S}}\|_{\ell_1} + 4\lambda_d \|\beta_{\mathcal{S}^c}^*\|_{\ell_1}.$$

Adding $(\lambda_d - \lambda_n^0) \|h_{\mathcal{S}}\|_{\ell_1}$ on both sides, we conclude the proof. \square

Using (6) and (28), it follows that

$$(31) \quad \frac{1}{4\lambda_\ell} \left[\|Xh\|_{\ell_2}^2 + (\lambda_\ell - \lambda_n^0) \|h\|_{\ell_1} \right] \leq \Delta\sqrt{s} \|Xh\|_{\ell_2} + \kappa_0 \|h\|_{\ell_1} + \|\beta_{\mathcal{S}^c}^*\|_{\ell_1}.$$

It yields,

$$\begin{aligned} \left[\frac{1}{4} \left(1 - \frac{\lambda_n^0}{\lambda_\ell} \right) - \kappa_0 \right] \|h\|_{\ell_1} &\leq \left(-\frac{1}{4\lambda_\ell} \|Xh\|_{\ell_2}^2 + \Delta\sqrt{s} \|Xh\|_{\ell_2} \right) + \|\beta_{\mathcal{S}^c}^*\|_{\ell_1}, \\ &\leq \lambda_\ell \Delta^2 s + \|\beta_{\mathcal{S}^c}^*\|_{\ell_1}, \end{aligned}$$

using the fact that the polynomial $x \mapsto -(1/4\lambda_\ell)x^2 + \Delta\sqrt{s}x$ is not greater than $\lambda_\ell \Delta^2 s$. This concludes the proof. \square

Proof of Theorem 1.4 and Theorem 1.6 — Using (27), we know that

$$\frac{1}{2\lambda_\ell} \left[\frac{1}{2} \|Xh\|_{\ell_2}^2 + (\lambda_\ell - \lambda_n^0) \|h\|_{\ell_1} \right] \leq \Delta\sqrt{s} \|Xh\|_{\ell_2} + \kappa_0 \|h\|_{\ell_1} + \|\beta_{\mathcal{S}^c}^*\|_{\ell_1}.$$

It follows that

$$\|Xh\|_{\ell_2}^2 - 4\lambda_\ell \Delta\sqrt{s} \|Xh\|_{\ell_2} \leq 4\lambda_\ell \|\beta_{\mathcal{S}^c}^*\|_{\ell_1}.$$

This latter is of the form $x^2 - bx \leq c$ which implies that $x \leq b + c/b$. Hence,

$$\|Xh\|_{\ell_2} \leq 4\lambda_\ell \Delta\sqrt{s} + \frac{\|\beta_{\mathcal{S}^c}^*\|_{\ell_1}}{\Delta\sqrt{s}}.$$

The same analysis holds for Theorem 1.6. \square

Acknowledgments — I would like to thank Jean-Marc Azaïs and Franck Barthe for their support.

REFERENCES

- [BRT09] P.J. Bickel, Y. Ritov, and A.B. Tsybakov, *Simultaneous analysis of lasso and Dantzig selector*, Ann. Statist. **37** (2009), no. 4, 1705–1732. MR 2533469 (2010j:62118)
- [CDD09] A. Cohen, W. Dahmen, and R. DeVore, *Compressed sensing and best k-term approximation*, J. Amer. Math. Soc. **22** (2009), no. 1, 211–231. MR 2449058 (2010d:94024)
- [CP09] E.J. Candès and Y. Plan, *Near-ideal model selection by ℓ_1 minimization*, Ann. Statist. **37** (2009), no. 5A, 2145–2177. MR 2543688 (2010j:62017)
- [CRT06] E.J. Candès, J.K. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math. **59** (2006), no. 8, 1207–1223. MR 2230846 (2007f:94007)
- [CT07] E.J. Candès and T. Tao, *The Dantzig selector: statistical estimation when p is much larger than n*, Ann. Statist. **35** (2007), no. 6, 2313–2351. MR 2382644 (2009b:62016)
- [DET06] D.L. Donoho, M. Elad, and V.N. Temlyakov, *Stable recovery of sparse overcomplete representations in the presence of noise*, IEEE Trans. Inform. Theory **52** (2006), no. 1, 6–18. MR 2237332 (2007d:94007)
- [DeV07] R.A. DeVore, *Deterministic constructions of compressed sensing matrices*, Journal of Complexity **23** (2007), no. 4-6, 918–925.
- [Don06] D.L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory **52** (2006), no. 4, 1289–1306. MR 2241189 (2007e:94013)
- [DS89] D.L. Donoho and P.B. Stark, *Uncertainty principles and signal recovery*, SIAM J. Appl. Math. **49** (1989), no. 3, 906–931. MR 997928 (90c:42003)
- [GLR08] V. Guruswami, J. R. Lee, and A. Razborov, *Almost euclidean subspaces of ℓ_1 via expander codes*, Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics, 2008, pp. 353–362.
- [Ind07] P. Indyk, *Uncertainty principles, extractors, and explicit embeddings of ℓ_2 into ℓ_1* , Proceedings of the thirty-ninth annual ACM symposium on Theory of computing, ACM, 2007, pp. 615–620.
- [IS10] P. Indyk and S. Szarek, *Almost-euclidean subspaces of ℓ_1 via tensor products: A simple approach to randomness reduction*, Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (2010), 632–641.
- [JN08] A. Juditsky and A. Nemirovski, *On verifiable sufficient conditions for sparse signal recovery via ℓ_1 minimization*, 2008.
- [JN10] ———, *Accuracy guarantees for ℓ_1 -recovery*.

- [Kas77] B.S. Kashin, *The widths of certain finite-dimensional sets and classes of smooth functions*, Izv. Akad. Nauk SSSR Ser. Mat. **41** (1977), no. 2, 334–351, 478. MR 0481792 (58 #1891)
- [KT07] B.S. Kashin and V.N. Temlyakov, *A remark on the problem of compressed sensing*, Mat. Zametki **82** (2007), no. 6, 829–837. MR 2399963 (2009k:41027)
- [Tib96] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Roy. Statist. Soc. Ser. B **58** (1996), no. 1, 267–288. MR 1379242 (96j:62134)
- [vdGB09] S.A. van de Geer and P. Bühlmann, *On the conditions used to prove oracle results for the Lasso*, Electron. J. Stat. **3** (2009), 1360–1392. MR 2576316
- [Š68] Z. Šidák, *On multivariate normal probabilities of rectangles: Their dependence on correlations*, Ann. Math. Statist. **39** (1968), 1425–1434. MR 0230403 (37 #5965)
- [ZY06] P. Zhao and B. Yu, *On model selection consistency of Lasso*, J. Mach. Learn. Res. **7** (2006), 2541–2563. MR 2274449

INSTITUT DE MATHÉMATIQUES DE TOULOUSE (CNRS UMR 5219), UNIVERSITÉ PAUL SABATIER, 118 ROUTE DE NARBONNE, 31062 TOULOUSE, FRANCE.

E-mail address: `yohann.decastro@math.univ-toulouse.fr`